



Erfassung und Entwicklung von Systemkompetenz – Empirische Befunde zu Kompetenzstruktur und Förderbarkeit durch den Einsatz analoger und digitaler Modelle

Measurement and Development of Systems Thinking–Empirical Findings Regarding the Structure of System Competence and Its Promotion by Means of Analogue and Digital Models

Medición y desarrollo del Pensamiento Sistémico: hallazgos empíricos sobre la estructura del sistema de competencias y su promoción mediante modelos analógicos y digitales

Svenja Brockmüller, Alexander Siegmund

Zusammenfassung

Die Herausforderungen im Kontext des globalen Wandels verdeutlichen die Relevanz von Systemkompetenz zur Beurteilung des Verhaltens komplexer dynamischer Mensch-Umwelt-Systeme. Diese empirische Vergleichsstudie untersucht im experimentellen Prä-Post-Test-Design die Veränderung der Systemkompetenz von Schülerinnen und Schülern am Themenbeispiel Bodenerosion durch die Arbeit mit (1) analogen bzw. (2) digitalen Modellen sowie (3) einer Kombination beider Zugänge. Ein auf Grundlage des heuristischen Freiburger Systemkompetenzstrukturmodells entwickelter Kompetenztest wird validiert und dessen vierdimensionale Struktur faktorenanalytisch fundiert. Varianzanalysen zeigen, dass der kombinierte Einsatz analoger und digitaler Modelle bei gleichem Zeitaufwand zu einer signifikant besseren Systemkompetenzförderung führt als der alleinige Einsatz digitaler Modelle.

Abstract

The challenges within the context of global climate change illustrate the relevance of system competence for evaluating complex human-environmental relations. By means of pre-post design this empirical comparison study examines the changes of the contextual systems thinking of students using (1) analogue or (2) digital soil erosion models (simulation), respectively (3) a combination of both approaches. Based on the "Freiburg heuristic competence model of systems thinking", test items are developed and validated. Confirmatory factor analyses suggest a four-dimensional competence structure. Analyses of variance reveal a significant higher mean value of system competence in the group that has experienced a combined analogue and digital intervention compared to the group with mere computer simulation.

Resumen

Los desafíos dentro del contexto del cambio climático global ponen de manifiesto la necesidad de un sistema de competencias que permita hacer una evaluación adecuada de las complejas relaciones entre el ser humano y el medioambiente. Por medio del diseño pre-posttest, el estudio de comparación empírica aquí presentado examina los cambios en el marco del pensamiento sistémico de los estudiantes (nivel secundario superior) usando (1) un modelo de erosión del suelo análogo, (2) un modelo digital de erosión del suelo (simulación) y (3) una combinación de ambos enfoques con la misma duración de cada intervención. Basado en el "Sistema de Pensamiento Heurístico de Freiburg" los ítems se desarrollaron y fueron validados. Los análisis factoriales confirmatorios (CFA) confirmaron una estructura cuatrodimensional de competencias. El análisis de varianza confirmó un valor medio de la competencia del sistema significativamente mayor en el grupo que ha experimentado una intervención basada en la combinación de medios en comparación con el grupo con mera simulación por computadora.

Zeitschrift für Geographiedidaktik
Journal of Geography Education
2020
Vol. 48(1), 31–49

© The author(s) 2020

Reprints and permission:

www.zgd-journal.de

ISSN 2698-6752

DOI <https://doi.org/10.18452/21389>



Schlüsselwörter

Systemkompetenz,
Interventionsstudie,
Testentwicklung,
Kompetenzstruktur,
Bodenerosionsmodelle

Keywords

system competence,
intervention study,
test development,
competence structure,
soil erosion models

Palabras clave

sistema de competencias,
estudio de intervención,
diseño de pruebas,
estructura de las
competencias,
modelos de erosión del
suelo

Autorin und Autor:

Dr. Svenja Brockmüller | Universität Koblenz-Landau | brockmueller@uni-landau.de

Prof. Dr. Alexander Siegmund | Pädagogische Hochschule Heidelberg | siegmund@ph-heidelberg.de

1 Begründung und Zielsetzung der Studie

Das Systemkonzept stellt eines der zentralen Basiskonzepte der Geographie dar: Es umfasst die Betrachtung der Erde als Mensch-Umwelt-System unter räumlicher Perspektive und die Analyse der Wechselbeziehungen zwischen den natur- und humangeographischen Subsystemen (DGfG, 2017, S. 10–12). Daher zählt es auch zu den zentralen Zielen des Geographieunterrichts, bei Schülerinnen und Schülern das hierfür notwendige systemische Denken zu fördern, indem Entwicklungen und Problemstellungen in Räumen untersucht werden, bei denen natur- und humangeographische Faktoren in ihrem Zusammenwirken betrachtet und analysiert werden können.

In der Fähigkeit zu systemischem Denken liegt der Schlüssel zu einem systemisch adäquaten und somit zukunftsorientierten Raumverhalten: Systemisches Denken ist hierbei als die Fähigkeit definiert, einen komplexen Wirklichkeitsbereich sozialer und/oder natürlicher Prägung unterschiedlicher Maßstabsgröße in seiner Struktur und seinem Verhalten als System zu erkennen, zu beschreiben, zu modellieren sowie auf Basis dieser Modellierung Prognosen und Maßnahmen zur Systemnutzung und -regulation zu treffen (REMPFLER & MEHREN, 2018, S. 205). Durch die funktionalen Verflechtungen der geosphärischen Welt auf jeder Maßstabsebene und in jedem Sachbereich mit der daraus resultierenden Ausbildung von Wechselwirkungssystemen fordert Köck (2018, S. 121) eine von Beginn an lernergerechte Herausarbeitung der systemischen Dimension im Geographieunterricht. Auch im Rahmen einer Bildung für eine nachhaltige Entwicklung (BNE) wird Systemverständnis als Grundlage für die Analyse von Wertmaßstäben auf ökonomischer, ökologischer, gesellschaftlicher und politischer Ebene verstanden (vgl. z.B. ROST, LAUSTRÖER & RAAK, 2003; OHL, 2018). RIESS (2013, S. 60) unterstreicht den Gedanken, dass Personen sich nur dann an einer umweltgerechten bzw. nachhaltigen Entwicklung von Räumen beteiligen können, wenn sie komplexe Zusammenhänge erkennen und verstehen, da sie erst dadurch in die Lage versetzt werden, beeinflussend einzugreifen. Der Systemwissenschaftler Hartmut BOSSEL (2004, S. 15) ergänzt zur Frage der Zukunftsrelevanz folgendermaßen: „Oft genug hängt das Leben von Individuen, die zukünftige Entwicklung einer Region oder gar die globale Zukunft ab von der Entwicklung dynamischer Systeme, die wir nicht ausreichend kennen, deren Verhalten wir nicht genug verstehen“.

Der Förderung von Systemkompetenz in räumlichen Kontexten kommt daher v.a. bei Kindern und Ju-

gendlichen als Akteuren von morgen eine maßgebliche Bedeutung zu. Dabei ergibt sich aus didaktischer Sicht die zentrale Forschungsfrage, wie – genauer: mit welchen Methoden – ein größtmöglicher Kompetenzzuwachs erreicht werden kann.

Zentrales Ziel der hier vorgestellten Studie (vgl. BROCKMÜLLER, 2019) ist es deshalb, unterschiedliche methodische Zugänge zur Förderung von Systemkompetenz empirisch miteinander zu vergleichen. Zielgruppe sind Schülerinnen und Schüler der gymnasialen Klassenstufe 10–12 in Baden-Württemberg. Aufbauend auf Themenfelder des Geographieunterrichts in der Mittelstufe zur Pedosphäre und Atmosphäre stellen die entwickelten Lerneinheiten dabei die Analyse raumrelevanter systemischer Strukturen und Prozesse einschließlich deren zukünftiger Entwicklung im Kontext der globalen Herausforderungen in den Vordergrund, wie sie für den Erwerb von Fachkompetenzen und damit verbundener fachspezifischer Methodenkompetenzen im Bildungsplan für die Oberstufe verankert sind. Exemplarisch ausgewählt wurde das Themenfeld der Bodenerosion, die mit ihren vielfältigen natürlichen und anthropogenen Einflussfaktoren ein idealtypisches Beispiel eines komplexen und dynamischen Mensch-Umwelt-Systems darstellt. Die Abschätzung von Erosionsfolgen sowie die Entwicklung geeigneter Anpassungsstrategien und Maßnahmen zum Schutz der Ressource Boden gelten dabei auch vor dem Hintergrund der regionalen Folgen des globalen Klimawandels als gesellschaftliche Herausforderung (vgl. FLAIG, 2013). Zur Förderung von Systemkompetenz wird das Themenbeispiel als geeignet angesehen, da es, den Empfehlungen von MEHREN, REMPFLE und ULLRICH-RIEDHAMMER (2014, S. 4–8) folgend, am konkreten geographischen Inhalt verankert ist, alle Dimensionen des zugrunde gelegten Systemkompetenzstrukturmodells adressiert und während der Erarbeitung graphische Repräsentationen des Systems in Form von Wirkdiagrammen sowie die entsprechenden Dynamiken visualisierende Modelle bzw. Computersimulationen zur Vermittlung der Systemeigenschaften zum Einsatz kommen (vgl. Kap. 4).

Als methodische Zugänge werden in der vorliegenden Studie ein analoges und ein digitales Bodenerosionsmodell gewählt, die im Rahmen der Treatments allein und in Kombination vergleichend eingesetzt werden. Mit ihrer Hilfe erarbeiten Schülerinnen und Schüler selbstständig hypothesenprüfend die systemischen Zusammenhänge im Rahmen eigens entwickelter Lernmodule. Während die grundsätzliche Eignung des didaktischen Einsatzes von Modellen

zur Systemkompetenzförderung vielfach theoriebasiert hergeleitet wurde (z.B. WIKTORIN, 2013; BROCKMÜLLER et al., 2016a), stellen sowohl eine empirische Absicherung der Förderwirksamkeit von Modellen, als insbesondere auch ein möglicher differenzierender Einfluss analoger bzw. digitaler Modelle Forschungsdesiderate dar, die im folgenden Kapitel 2.1 genauer dargelegt werden.

Zur Erfassung von Veränderungen der Systemkompetenz wird ein kontextspezifisches Messinstrument für einen Prä-Post-Einsatz entwickelt und validiert. Mit Hilfe dieses Instruments wird untersucht, ob verschiedene methodische Zugänge – namentlich der Einsatz analoger, digitaler oder kombinierter Modelle – das kontextuelle systemische Denken von Schü-

lerinnen und Schülern generell bzw. in den einzelnen Teildimensionen unterschiedlich fördern.

Ein weiteres Ziel der Studie ist, einen Beitrag zur Entschlüsselung der Struktur des Konstrukts Systemkompetenz und seiner Dimensionalität zu leisten, da hierzu in der geographiedidaktischen Forschung heterogene Annahmen und Befunde vorliegen. Besonderes Forschungsinteresse liegt auf der Frage, ob das vierdimensionale Freiburger heuristische Kompetenzstrukturmodell zum systemischen Denken, das von Arbeitsgruppe um Werner Rieß entwickelt wurde (RIESS, SCHULER & HÖRSCH, 2015), im Kontext und anhand der Daten der vorliegenden Studie empirisch gestützt werden kann.

2 Theoretischer Hintergrund und Forschungsdesiderate

Entsprechend der beiden hergeleiteten Untersuchungsbereiche Systemkompetenzförderung durch Modelleinsatz bzw. Dimensionalität von Systemkompetenz werden im Folgenden zentrale theoretische Grundlagen der Studie und die darauf aufbauend adressierten Forschungsdesiderate zusammenfassend dargestellt.

2.1 Analoge und digitale Modelle im didaktischen Vergleich

Didaktische Stärken von Modellen sind die Visualisierung und Vereinfachung komplexer Systeme oder Prozesse sowie deren akzentuierte Strukturierung (WIKTORIN, 2014, S. 6–8). BIRKENHAUER (1997) empfiehlt den Einsatz von Modellen zum Verständnis und zur Reflexion räumlicher Fragestellungen sowie zur Evaluation verschiedener Lösungsmöglichkeiten im Sinne einer verantwortungsvollen Raumplanung. Modelle, zur Ergebniserwartung eingesetzt oder selbst das Ergebnis von kognitiven Prozessen, eignen sich, um Erklärungen zu formulieren oder Vorhersagen zu treffen (SCHUBERT, 2013). Hierbei können insbesondere Modelle, die nicht nur die Struktur oder Funktion von Objekten visualisieren, sondern auch dynamische Prozesse mit ihren möglichen Rückkopplungen und nicht-linearen Abhängigkeiten enthüllen, die verschiedenen Dimensionen und Stufen von Systemkompetenz adressieren (vgl. BROCKMÜLLER et al., 2016a, S. 160). In Bezug auf eine Förderwirksamkeit von Systemkompetenz durch Modelle leiten BROCKMÜLLER et al. (2016a, S. 166) Modell-Designkriterien in Bezug auf Konstruktion, Anwendung und Instruktion zunächst theoretisch her. Dies sind beispielsweise

eine unabhängige Manipulationsmöglichkeit individueller Systemparameter, eine Sichtbarkeit aller im Sinne des Modellzwecks relevanten Prozesse oder eine instruktionale Fokussierung auf die Interaktionen von Systemparametern. Bezogen auf die Funktionen hypothesenprüfender Forschungsexperimente nach OTTO (2015, S. 146) können Schülerinnen und Schüler an analogen oder digitalen Modellen einen Prozess skaliert oder zeitgerafft nachahmen, dabei die Wirkung einzelner Faktoren gezielt verändern und Vorstellungen über oder Einsichten in Prozessverläufe gewinnen.

Während analoge Modelle hierbei Informationen über haptische und visuelle Hinweise bereitstellen und bestimmte Handlungen, z.B. das Zufügen von Substanzen, erfordern, liefern digitale Modelle diese Informationen entsprechend ausschließlich visuell und werden per Tastatur oder Maus bedient (vgl. RUDDLE & JONES, 2001; TRIONA & KLAHR, 2003). Köck (2018, S. 120) weist darauf hin, dass per *Knopfdruck* IT-gestützt ausgeführte Operationen ggf. zu strukturell-funktionalem Nicht-Verstehen führen können. Aus der Bedienung von komplexen und ungewohnten Simulationsumgebung kann ggf. eine kognitive Überlastung der Lernenden resultieren (vgl. PAAS, TUOVINEN, TABBERS & VAN GERVEN, 2003). Als besondere Stärke von Computersimulationen wird hingegen vor allem die Modellierbarkeit zeitlicher Veränderungen sowie die Erfassung prozessual-dynamischer Eigenschaften gesehen (z.B. BOLLMANN-ZUBERBÜHLER & KUNZ, 2008, S. 36; EDSALL & WENTZ, 2007, S. 428; RIESS, 2013; JEBBINK, 2013, S. 254–255; KÖCK, 2018, S. 115). Durch die Kombination von Modell und Computersimulation ist, besonders bei komplexen Sachverhal-

ten, ein vertieftes, nuanciertes Verständnis möglich (DE JONG, LINN & ZACHARIAS, 2013; WINBERG & BERG, 2007).

Vorliegende Studien zum unterrichtlichen Einsatz von Modellen/Simulationen beziehen sich jedoch meist nicht explizit auf die Förderung von Systemkompetenz, darüber hinaus wird häufig der Einsatz analoger oder digitaler Modelle gegen eine Kontrollgruppe ohne Modelleinsatz verglichen, nicht jedoch der Einfluss eines kombinierten Modelleinsatzes überprüft. EDSALL und WENTZ (2007) stellen beispielsweise in ihrer Studie zur Untersuchung der Wirksamkeit von computerbasierten Anwendungen in der geographischen Bildung analoge und digitale Kartenprojektionen vergleichend gegenüber, welche sich ihren Ergebnissen zufolge in ihrer Lernwirksamkeit nicht signifikant voneinander unterscheiden. Ohne die Kombination beider Methoden in der eigenen empirischen Studie untersucht zu haben, vertreten EDSALL und WENTZ (2007, S. 429) abschließend die These, dass diese Verbindung der analogen und digitalen Methode wohl die beste Lösung sein müsse, Lehrkräfte jedoch aufgrund ihrer beschränkten v.a. zeitlichen Ressourcen Unterstützung bei der Entscheidung für *die eine beste Methode* benötigen.

STULL und HEGARTY (2016) untersuchen im Kontext der Chemiedidaktik die Förderung von Darstellungskompetenz durch den Einsatz dreidimensionaler digitaler bzw. gegenständlicher Modelle gegenüber einer Kontrollgruppe ohne Modelleinsatz, zwischen den digitalen und gegenständlichen Modellen zeigt sich jedoch kein statistisch bedeutsamer Unterschied. STULL und HEGARTY (2016, S. 15) schlussfolgern, dass der Einsatz von manipulierbaren, interaktiven Modellen vielfältige Lernprozesse generell fördere, wobei es nicht darauf ankomme, ob es sich um digitale oder gegenständliche Modelle handele. Zudem erfordern digitale Modelle mehr Zeit für gleich gute Lösungen als analoge Modelle (STULL & HEGARTY, 2016, S. 8).

In der Mathematikdidaktik kommen LICHTI und ROTH (2018) in der Beforschung der Förderung funktionalen Denkens bei Schülerinnen und Schülern der Klassenstufe 6 zu dem Ergebnis, dass sowohl Computersimulationen als auch gegenständliche Materialien – im Gegensatz zur Kontrollgruppe – das funktionale Denken signifikant fördern und im direkten Vergleich von Computersimulation und gegenständlichen Materialien eine signifikant größere Steigerung des funktionalen Denkens durch Einsatz der Computersimulationen erreicht wird (LICHTI & ROTH, 2018, S. 96). Den nicht untersuchten kombinierten Einsatz beider Medien empfehlen LICHTI und ROTH (2018, S. 96) abschließend für den Schulunterricht (LICHTI & ROTH,

2018, S. 101), um das verbesserte qualitative Verständnis funktionaler Zusammenhänge durch Simulationen mit dem verbesserten quantitativen Verständnis durch gegenständliche Materialien zu verknüpfen.

In der Biologiedidaktik untersuchen RIESS und MISCHO (2008) die Wirkung variierten Unterrichts auf systemisches Denken von Schülerinnen und Schülern der Klassenstufe 6 mittels Computersimulation (4h), *klassischer* Unterrichtseinheit (11h) sowie Kombination aus beidem (2h + 9h). Sie stellen bei der kombinierten Gruppe den höchsten Lernzuwachs bei jedoch nicht signifikanten Gruppenunterschieden fest. Aufgrund der sehr unterschiedlichen Zeitdauer der drei vergleichend untersuchten Interventionen ist die Aussagekraft dieser Ergebnisse jedoch eingeschränkt.

Vor diesem Hintergrund ergibt sich ein Forschungsdesiderat zur Frage, ob der Einsatz analoger, digitaler oder kombiniert eingesetzter Modelle bei gleicher Zeitdauer der jeweiligen Interventionen die Systemkompetenz von Schülerinnen und Schülern am effektivsten fördert. Die Klärung, hier exemplarisch im geographischen Kontext des Teilsystems Bodenerosion, weist dabei bildungstheoretische sowie unterrichtspraktische Relevanz im Kontext der Naturwissenschaftsdidaktiken auf.

2.2 Systemkompetenzstruktur

Betrachtet man Systemkompetenz im Sinne von ROST et al. (2003), SENGE und KLOSTERMANN (2006), SOMMER (2006) oder RIESS und MISCHO (2008) als Beitrag zur Befähigung von Schülerinnen und Schülern, aktiv und verantwortungsvoll zu einer nachhaltigen Entwicklung beizutragen, muss hierbei insbesondere auf das Denken in komplexen Zusammenhängen (vgl. MEHREN, MEHREN, OHL & RESENBERGER, 2015, S. 4) fokussiert werden, in einer Weise, die zur Bildung fachlich fundierter Urteile führt und keine monokausalen Lösungsansätze suggeriert (MEHREN, 2018, S. 134; OHL, 2018, S. 121).

Zum Konstrukt Systemkompetenz und seiner Dimensionalität liegen heterogene Annahmen und Befunde vor, so geht z.B. BRÄUTIGAM (2014) zunächst von Eindimensionalität aus, während VIEHRIG et al. (2012, 2017) ein geographiedidaktisch begründetes Kompetenzstruktur- und -stufenmodell mit den drei Dimensionen 1: *Systeme erfassen und analysieren*, 2: *Handlungsmöglichkeiten gegenüber Systemen beurteilen* und 3: *Räumliches Denken* aufstellen und untersuchen. Da sich diese Dimensionen psychometrisch jedoch nicht separat absichern lassen, schließen sie auf eine hohe Korrelation systemischen und räumlichen Denkens in geographischen Kontexten.

MEHREN et al. (2015, 2016, 2018) gelingt es schließlich, ein Kompetenzstruktur- und -stufenmodell zur Erfassung geographischer Systemkompetenz mit den beiden Dimensionen 1: *Systemorganisation und -verhalten* sowie 2: *Systemadäquate Handlungsintention* empirisch zu validieren. Hierbei ist das räumliche Denken als immanenter Teil erdräumlicher Systeme implizit einbezogen (MEHREN et al., 2016, S. 154).

Von der interdisziplinären Arbeitsgruppe um Rieß wurde zeitgleich das vierdimensionale Freiburger heuristische Kompetenzstrukturmodell zum systemischen Denken entwickelt, in welchem folgende vier Dimensionen ausgewiesen werden (RIESS et al., 2015, S. 19).

A: Deklaratives/konzeptuelles Systemisches Wissen

Diese Dimension fokussiert sowohl auf grundlegende kognitive Komponenten systemischen Denkens, wie z.B. Grundsätze einer systemwissenschaftlichen Betrachtungsweise, als auch auf Eigenschaften komplexer Systeme, wie z. B. Nicht-Linearität oder Emergenz.

B: Systemmodellierungsfähigkeit

Diese Dimension adressiert insbesondere solche Fähigkeiten, die sich auf Methoden der Systemwissenschaften beziehen, wie z.B. Verständnis und Konstruktion unterschiedlicher qualitativer und quantitativer Systemmodelle.

C: Fähigkeit zur Nutzung von Systemmodellen beim Lösen von komplexen dynamischen Problemen

Diese Dimension zielt auf systemische Problemlösefähigkeit ab, z.B. das Treffen von Entscheidungen über die Verwendbarkeit von Systemmodellen zu Erklärungs- oder Prognosezwecken inklusive einer Beurteilung der Reichweite der auf Basis der Modellierung gewonnen Erkenntnisse.

D: Bewertung von Systemmodellen und Ergebnissen der Modellanwendung

Diese Dimension nimmt schließlich reflexive Fähigkeiten in den Blick, wie z.B. die Bestimmung der Struktur-, Verhaltens- und Anwendungsgültigkeit sowie der

Vorhersageunsicherheit von Systemmodellen bzw. Ergebnissen der Modellanwendung.

Die Kurzbeschreibung der Dimensionen lässt erkennen, dass beim Freiburger Kompetenzstrukturmodell ein besonderer Fokus auf der Konstruktion und Anwendung von Systemmodellen liegt. Damit weist es eine besondere Passung zur Forschungsfrage der vorliegenden Studie auf, welche die Förderbarkeit von Systemkompetenz durch Einsatz unterschiedlicher (analoger bzw. digitaler) Bodenerosionsmodelle in den Blick nimmt. Das heuristische Freiburger Systemkompetenzstrukturmodell wird daher als maßgebende theoretische Rahmung zugrunde gelegt. Die Faktorenstruktur dieses Modells wurde bislang nicht anhand empirischer Daten überprüft. Es wird daher im Rahmen der hier vorgestellten Studie zunächst faktorenanalytisch untersucht, ob die anhand des Fallbeispiels Bodenerosion erhobenen Daten tatsächlich die theoretisch postulierten vier Dimensionen abbilden. Im Vergleich der drei vorgestellten Modelle fällt auf, dass die hier beschriebene Dimension B *Systemmodellierungsfähigkeit* (RIESS et al., 2015) mit den beiden oben genannten Dimensionen 1: *Systemorganisation und -verhalten* von MEHREN et al. (2015) bzw. *Systeme erfassen und analysieren* von VIEHRIG et al. (2017) vergleichbar ist, während Dimension C *Systemmodellnutzung* Parallelen zu den beiden Dimensionen 2: *Systemadäquate Handlungsintention* von MEHREN et al. (2015) bzw. *Handlungsmöglichkeiten gegenüber Systemen beurteilen* von VIEHRIG et al. (2017) aufweist. Diese Entsprechungen wurde auch in BROCKMÜLLER, VOLZ und SIEGMUND (2016b, S. 110) sowie MEHREN et al. (2016, S. 153) bereits beschrieben und begründet. Resultierend ergibt sich insbesondere die Frage, ob die beiden von RIESS et al. (2015) zusätzlich ausgewiesenen Dimensionen A *Konzeptuelles systemisches Wissen* und D *Bewertung von Systemmodellen* im Untersuchungskontext eigene Dimensionen des Konstrukts abbilden oder – so die Alternative – Facetten oder Schwierigkeitsstufen der beiden zentralen Dimensionen B und C darstellen und somit implizit in diesen enthalten sind (vgl. MEHREN et al., 2016, S. 154).

3 Studiendesign und Forschungsmethodik

Aus dem dargestellten Forschungsstand ergeben sich zwei Teilstudien mit aufeinander aufbauenden Aufgaben und Forschungsfragen: In der ersten Teilstudie wird zunächst ein Systemkompetenztest im Kontext raumwirksamer Mensch-Umwelt-Beziehungen (Bodenerosion) entwickelt. Zu prüfen ist hier, ob der ent-

wickelte Test Systemkompetenz valide, reliabel und objektiv misst. Teil dieser Analysen ist außerdem die Frage, ob das theoretisch als vierdimensional angenommene Konstrukt Systemkompetenz mit Hilfe der beobachteten Daten faktorenanalytisch belegt werden kann. Hierzu werden, nach einer vorangestellten

exploratorischen Faktorenanalyse, konfirmatorische Faktorenanalysen erster und zweiter Ordnung mit Hilfe des Paketes *lavaan* der Statistiksoftware R (ROSSEEL, 2012) durchgeführt und mittels des Paketes *semPlot* (EPSKAMP & STUBER, 2017) graphisch dargestellt.

Im zweiten Teil der Studie wird der entwickelte und validierte Test schließlich als Prä-Post-Messinstrument im Rahmen einer empirischen Interventionsstudie eingesetzt: Hier erfassen, analysieren und beurteilen insgesamt 370 Schülerinnen und Schüler (Vorstudie I: 89, Pilotierungsstudie: 78, Hauptstudie: 203) der gymnasialen Klassenstufen 10 bis 12 die komplexen Einflussfaktoren der Bodenerosion. Aus der Forschungsfrage *Welcher methodische Zugang ist zur Förderung von geographischer Systemkompetenz am effektivsten?* ergibt sich dabei eine Einteilung der Schülerinnen und Schüler in drei Gruppen mit unterschiedlichen methodischen Zugängen:

- Treatment-Gruppe 1: Einsatz gegenständlicher Bodenerosionsmodelle (*analoges Modell*)
- Treatment-Gruppe 2: Einsatz von Computersimulationen zum Bodenabtrag (*digitales Modell*)
- Treatment-Gruppe 3: kombinierter Einsatz von analogem und digitalem Modell (*Kombination*)

Im Prä-Post-Test-Design wird im beschriebenen Kontext untersucht, welcher methodische Zugang (= Treatmentgruppe, vgl. ABB. 1) die Systemkompetenz von Schülerinnen und Schülern bzw. deren Teilaspekte/Dimensionen am wirksamsten fördert. Die Systemkompetenz der Schülerinnen und Schüler wird damit im Vorher-Nachher-Vergleich metrisch als abhängige Variable erfasst, die drei Treatmentgruppen dienen

zur Erklärung möglicher Unterschiede der abhängigen Variablen.

Hypothetische Annahmen hierzu sind, dass die aktive Herangehensweise (hypothesenprüfendes Experimentieren) bei allen drei methodischen Zugängen Systemkompetenz bedeutsam fördert und die Kombination aus analogem Modell und Computersimulation bei gleicher Interventionsdauer den größten Systemkompetenzzuwachs erzielt.

Der erforderliche Stichprobenumfang der Hauptstudie wurde mittels G*Power (Version 3.1.9.2; FAUL, ERDFELDER, LANG & BUCHNER, 2007) ermittelt: Bei drei Gruppen und entsprechend zwei Zählerfreiheitsgraden, einer angenommenen mittleren Effektgröße f (0,25), einer α -Fehlerwahrscheinlichkeit von 5% sowie einer β -Fehlerwahrscheinlichkeit von 10% (= Teststärke von 90%) ergibt sich für das zugrunde gelegte Testdesign ein benötigter Umfang von 206 Testpersonen. Kleinere Abweichungen vom geplanten balancierten Design haben sich im Nachhinein durch Ausfall einzelner Testpersonen ergeben.

Die Auswertung erfolgte varianzanalytisch. Varianzanalysen (*analyses of variance*, ANOVA) dienen dabei dem Vergleich von mehr als zwei Gruppen hinsichtlich der zentralen Tendenz einer Variablen (hier: Systemkompetenz) ohne α -Fehler-Kumulierung. Das Vorgehen dient zur Klärung der Frage, ob die Variation der Mittelwerte *nur* den Stichprobenfehler widerspiegelt, also zufallsbedingt ist, oder ob die experimentelle Manipulation (hier: *Gruppe* – analoges Modell, digitales Modell, Kombination) einen Effekt hat, der statistisch signifikant ist.



ABB. 1 Treatmentgruppen (1) analoges Modell, (2) digitales Modell und (3) Kombination aus analogem und digitalem Modell der Interventionsstudie (Quelle: AUTORIN & AUTOR)

4 Entwicklung und Evaluation der Treatments und Testinstrumente

Der Fokus bei der Entwicklung der Lerneinheiten zum Thema Bodenerosion lag auf der Möglichkeit zur vernetzenden Betrachtung von Zusammenhängen und Rückkopplungen sowie der Beurteilung von Folgen verschiedener Schutz- und Anpassungsstrategien. Die angebotenen Lerngelegenheiten sollten grundsätzlich geeignet sein, Wissen zu verknüpfen und den Lerngegenstand in Relation zu weiteren Systemelementen zu setzen bzw. die Lernenden anregen, verschiedene Perspektiven einzunehmen sowie Informationen zu einem komplexeren Gesamtbild zusammenzuführen.

In einer Vorstudie mit 89 Schülerinnen und Schülern wurden die Treatments in sechs Design-Zyklen entwickelt, evaluiert und optimiert (vgl. BROCKMÜLLER, 2019, S. 252–270). Die Überarbeitung basierte hierbei v.a. auf Eigen- und Fremdbeobachtungen sowie Experten-Feedbacks insbesondere zu der Frage, ob die vier Teilbereiche der Systemkompetenz, wie RIESS et al. (2016) sie ausgewiesen haben, explizit adressiert sind. Die generelle Eignung der drei Treatments zur Förderung von Systemkompetenz im Sinne der Zielsetzung, nämlich konzeptuelles systemisches Wissen, Systemmodellierungsfähigkeit, systemische Problemlösefähigkeit sowie Systemmodellreflexion zu fördern, wird somit zunächst postuliert. Alle drei Treatments weisen die gleichen Aufgabenstellungen und eine identische Bearbeitungsdauer auf. Sie unterscheiden sich lediglich hinsichtlich der Arbeitsmethodik im Sinne der zur Verfügung stehenden analogen, digitalen oder kombinierten Modelle, an welchen jeweils hypothesenprüfend experimentiert wird. Hierbei wurde bei der Eigenentwicklung des analogen Modells (vgl. [ABB. 1](#)) darauf geachtet, dass es mit dem auf der Allgemeinen Bodenabtragsgleichung (ABAG) basierenden digitalen Modell (BRANDHUBER et al., 2012, Version 1.0, inzwischen ersetzt durch Version 2.0, 2018) vergleichbare Möglichkeiten zur Manipulation von Parametern aufweist.

Zur Überprüfung der Förderwirkung der Treatments konnte ein Systemkompetenztest im Kontext Bodenerosion in zwei Vorstudien mit insgesamt 45 Erstsemesterstudierenden sowie einer Pilotierungsstudie mit 78 Schülerinnen und Schülern basierend auf den Annahmen der klassischen Testtheorie entwickelt und überprüft werden (vgl. BROCKMÜLLER, 2019, S. 224–251). Das Testinstrument ist demnach geeig-

net, kontextuelles systemisches Denken von Schülerinnen und Schülern objektiv, valide, reliabel und ökonomisch zu erfassen. Die Itemschwierigkeiten (nach DAHL, 1971) liegen zwischen 12,9 und 84,0. Anhand der von zwei Ratern mit Hilfe einer Auswertungsmatrix bewerteten Tests (Anteil Doppelbegutachtung: 8,3%) wurde in Bezug auf die Auswertungsobjektivität eine sehr gute Interrater-Übereinstimmung mit einem Korrelationskoeffizienten von Pearson's $r = 0,964$ erzielt. Die Inhaltsvalidität des entwickelten Testes konnte auf Basis der Zuordnung der Testitems zu den Systemkompetenzdimensionen des Freiburger Systemkompetenzstrukturmodells mit einer substanziellen Übereinstimmung von sieben Experten-Ratings (Fleiss' $\kappa = 0,729$) belegt werden.

Die Reliabilität (Messgenauigkeit) wurde mithilfe des Kriteriums der internen Konsistenz ermittelt: Hierbei weisen die beiden Items der Dimension A einen Cronbachs α -Wert von 0,54 auf, die fünf bzw. vier Items der Dimensionen B und D erreichen jeweils den Wert 0,78 und die drei Items der Dimension C den Wert 0,70. Nach FISSEN (1997) interpretiert weisen Skalen erst ab 0,80 eine mittlere bzw. ab 0,90 eine hohe Reliabilität auf. GEORGE und MALLERY (2003) sprechen bereits ab 0,70 von einer akzeptablen Reliabilität, welche immerhin für drei der vier hier abgebildeten Skalen (Dimension B, C und D) zutrifft. Laut LIENERT und RAATZ (1998) soll ein Testinstrument zum Gruppenvergleich eine interne Konsistenz von mindestens 0,50 aufweisen, damit valide Aussagen zur Wirksamkeit einer Intervention getroffen werden können. Dieser kritische Wert wird in allen vier Testskalen überschritten, sodass der entwickelte Test als geeignetes Instrument zur Beantwortung der Forschungsfragen der Interventionsstudie bewertet werden kann, während beim Vergleich mit weiteren Systemkompetenzmodellen insbesondere Dimension A, die aufgrund ihrer geringen Item-Ausstattung eine vergleichsweise schwache Reliabilität aufweist, nur mit besonderer Vorsicht interpretiert werden kann.

SEYBOLD, FRISCHKNECHT-TOBLER und NAGEL (2008) konstatieren Forschungsdesiderate im Kontext Systemdenken, welche u.a. in der Ausarbeitung von Messverfahren und der Entwicklung von Unterrichtsmaterial liegen. Die im Rahmen der hier vorgestellten Studie entwickelten Treatments und Instrumente leisten hierzu einen Beitrag.

5 Faktorenstruktur des zu Grunde gelegten Systemkompetenzmodells

Die Unsicherheiten um das Konstrukt *Systemkompetenz* und die unterschiedlichen Forschungsannahmen oder auch empirischen Ergebnisse bezüglich seiner Dimensionalität (vgl. Kap. 2.2) lassen es sinnvoll erscheinen, die Faktorenstruktur zunächst zu *explorieren*, bevor dann ein plausibles Modell nach Möglichkeit *konfirmiert* wird. Die durchgeführte exploratorische Faktorenanalyse (Maximum-Likelihood-Methode) lässt auf vier Faktoren (Eigenwerte 4,13 bzw. 0,97 bzw. 0,53 bzw. 0,38) schließen, auch die Parallelanalyse nach [HORN \(1965\)](#) stützt diese Annahme (vgl. [BROCKMÜLLER, 2019, S. 76–77](#)). Das theoretisch zugrunde gelegte vierdimensionale Freiburger Systemkompetenzstrukturmodell wird im nächsten Schritt mittels konfirmatorischer Faktorenanalyse auf seine

empirische Passung mit den Daten der vorliegenden Studie (Stichprobe der Hauptstudie, $n = 203$, Prätest) untersucht und mit alternativen Modellen verglichen.

5.1 Konfirmatorische Faktorenanalyse erster Ordnung

Um die Modellparameter zu schätzen und die konkurrierenden Modelle mit ein-, zwei-, drei- bzw. vierfaktorieller Lösung miteinander zu vergleichen, kommt die Schätzmethode Maximum-Likelihood zur Anwendung. Im Vergleich zeigt die vierfaktorielle Lösung anhand der folgend dargestellten Modellgütekoeffizienten die beste Modellpassung (vgl. [TAB. 1](#)).

TAB. 1: Modellvergleich mittels Modellgütekoeffizienten (Quelle: AUTORIN & AUTOR)

Faktoren	Freiheitsgrade	Gesamtanpassung des Modells		Modellvergleiche					
				Likelihood-Ratio-Differenzen-Test			Informationskriterien und Incremental-Fit-Indizes		
		SRMR	RMSEA	χ^2	χ^2 diff.	Diff. sign.?	AIC	BIC	CFI
1	76	0,097	0,144	400,61			11.781	11.810	0,653
2	77	0,083	0,098	223,55	177,07	***	11.543	11.639	0,842
3	74	0,068	0,077	163,97	59,57	***	11.487	11.590	0,904
4	71	0,061	0,069	138,80	25,18	***	11.468	11.580	0,927

Der SRMR (Standardized Root Mean Square Residual) ist ein Maß für das durchschnittliche Residuum ([EID, GOLLWITZER & SCHMITT, 2011, S. 880](#)). Werte nahe Null deuten auf eine gute Modellpassung hin. Eine bedeutsame Abweichung des Modells von den Daten liegt laut [HU und BENTLER \(1999\)](#) vor, wenn der SRMR 0,08 übersteigt. Der RMSEA (Root Mean Square Error of Approximation) zählt zu den Closeness/Badness-of-Fit-Koeffizienten. Der RMSEA stellt die durchschnittliche Abweichung der Daten vom Modell pro Freiheitsgrad dar, deswegen steht ein höherer Wert für eine schlechtere Modellpassung ([BÜHNER, 2011, S. 424–425](#)). [STEIGER \(2007, S. 893–898\)](#) setzt eine strikte RMSEA-Obergrenze bei 0,07 an, die lediglich von der vierfaktoriellen Lösung unterschritten wird ([TAB. 1](#)). Im direkten Modellvergleich der ein- bis vierfaktoriellen Lösung kann der Likelihood-Ratio-Differenzen-Test ([EID et al., 2011, S. 882](#)) angewendet werden, da die Modelle ineinander verschachtelt sind. Ob die Unterschiede der χ^2 -Werte statistisch bedeutsam (signifikant) sind, wird mittels der Differenz der

χ^2 -Werte als Prüfgröße ermittelt: Das jeweils weniger restriktive, also höher-faktorielle Modell passt dabei signifikant besser zu den Daten (vgl. [TAB. 1](#)).

Informationstheoretische Maße wie AIC (Akaike-Information-Criterion) oder BIC (Bayesian-Information-Criterion) eignen sich zum Vergleich mehrere Modelle, auch falls diese nicht ineinander geschachtelt sind (vgl. [EID et al., 2011, S. 883](#)). Diese Maße berücksichtigen neben der Modellgüte zusätzlich die Modellsparsamkeit. Da die hier herangezogenen Maße als die minimale Varianz der Residuen formuliert sind, ist das Modell mit dem niedrigsten Wert des jeweiligen Informationskriteriums am besten: Bei den vorliegenden Daten ([TAB. 1](#)) sprechen in dieser Hinsicht sowohl AIC als auch BIC für das vierdimensionale Modell. Bei Incremental-Fit-Indizes, z.B. dem CFI (Comparative-Fit-Index), wird die Anpassungsgüte eines Modells mit einem *Unabhängigkeits-* oder *Baseline-Modell* verglichen ([Bühner, 2011, S. 427](#)). Von [HU und BENTLER \(1999\)](#) wird ein CFI größer 0,90 als akzeptabel, größer 0,95 als gut bezeichnet. Das ein-

und zweifaktorielle Modell (TAB. 1) sind folglich nicht akzeptabel und müssen daher im Kontext der vorliegenden Studie verworfen werden.

Für die durchgeführten konfirmatorischen Faktorenanalysen erster Ordnung wurde jeweils ein Messmodell zur Zuordnung von manifesten Variablen (Items) als Indikatoren latenter Variablen (Systemkompetenzdimensionen) definiert. Das hier favorisierte vierdimensionale Messmodell ist im Folgenden graphisch als Pfaddiagramm mit unstandardisierten Modellparametern dargestellt (ABB. 2): Die manifesten Variablen sind durch Quadrate und die latenten Variablen durch Kreise gekennzeichnet. Hierbei geben Einfach-

die jeweilige Dimension einbezogenen Items immer 1 gesetzt (gestrichelte Pfeile) und die weiteren Faktorladungen werden jeweils relativ hierzu ausgegeben (durchgezogene Pfeile). Zudem zeigt ABB. 2 auch die Kovarianzen der vier Systemkompetenzdimensionen. Bei diesen ist insbesondere der relativ geringere Zusammenhang der Dimension A mit allen weiteren Dimensionen inhaltlich auffällig, während die Kovarianzen zwischen den drei Dimensionen B, C und D insgesamt höher ausfallen. Mögliche Ursachen hierfür werden in Kapitel 5.3 diskutiert. Im Vorgriff auf das folgende Kapitel 5.2 dient das hier aufgestellte Modell erster Ordnung (ABB. 2), das bisher lediglich Zu-

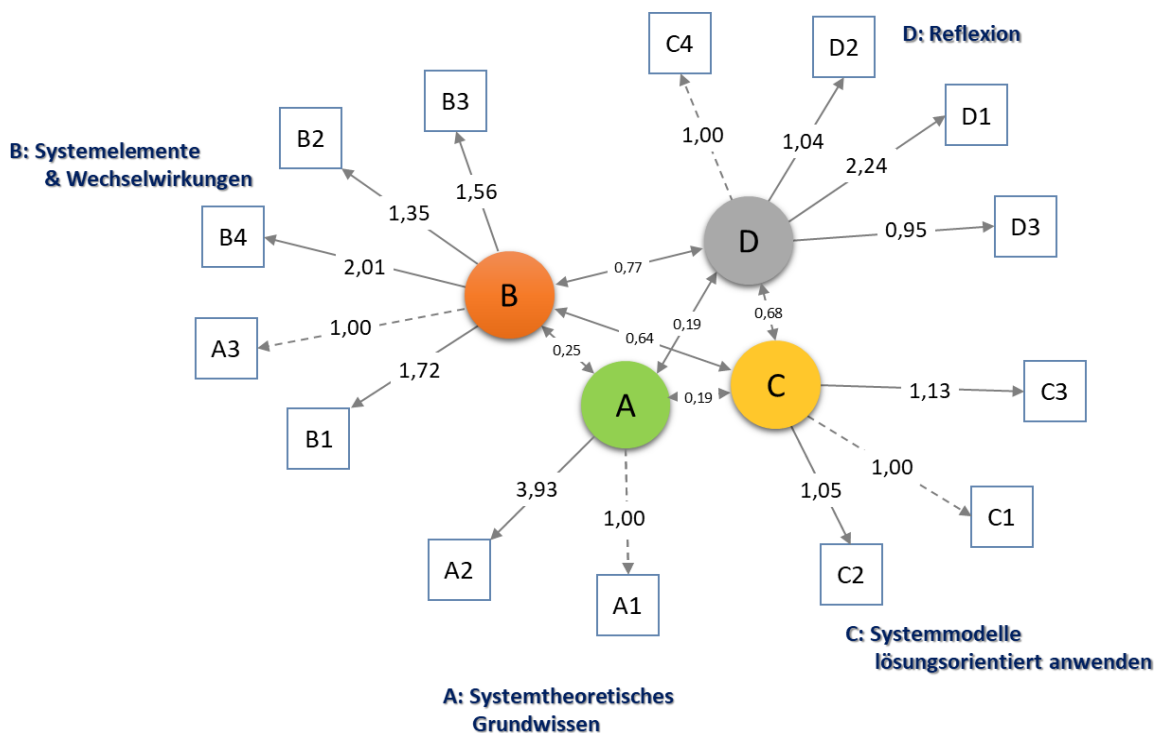


ABB. 2 Pfaddiagramm des Messmodells zur Systemkompetenz (First Order Model) (Quelle: AUTORIN & AUTOR)

pfeile die Stärke der gerichteten Beziehungen zwischen den latenten Variablen und den jeweiligen Indikatoren im Messmodell (auch: Ladungskoeffizienten) bzw. zwischen latenten Variablen im Strukturmodell (auch: Pfadkoeffizienten) an (BÜHNER, 2011, S. 384–385; SCHUMACKER & LOMAX, 2004). Je größer der Wert des jeweiligen Koeffizienten ist, desto größer ist die Bedeutung des jeweiligen Merkmals für das untersuchte Zielmerkmal (FREIN, MÖLLER, PETERMANN & WILPRICHT, 2007, S. 286). Ungerichtete Beziehungen (Kovarianzen, Korrelationen) zwischen den latenten Variablen sind durch Doppelpfeile dargestellt.

ABB. 2 zeigt die Faktorladungen der Items auf die jeweiligen Systemkompetenzdimensionen A bis D – dies ist die Varianz am Item, die durch den Faktor aufgeklärt wird. Hierbei ist die Faktorladung des ersten in

sammenhänge zwischen den Dimensionen abbildet, insbesondere dem statistischen Vergleich mit einem Modell zweiter Ordnung (ABB. 3), das den Beitrag der vier Dimensionen zum gemeinsamen übergeordneten Konstrukt Systemkompetenz berücksichtigt.

5.2 Konfirmatorische Faktorenanalyse zweiter Ordnung

Die konfirmatorische Faktorenanalyse zweiter Ordnung ist eine statistische Methode zur Bestätigung, dass ein theoretisch angenommenes Konstrukt auf eine bestimmte Anzahl untergeordneter Konstrukte (Subkonstrukte) lädt, beziehungsweise um herauszufinden wie stark der Effekt des Gesamtkonstrukts (jetzt: Konstrukt zweiter Ordnung, Second Order con-

struct) auf die einzelnen Subkonstrukte (jetzt: Konstrukte erster Ordnung, First Order constructs) ist. Hierzu können bei diesem Verfahren zusätzlich Faktorladungen zweiter Ordnung durch eine Technik der Faktorenanalyse direkt aus der Matrix der Korrelationen zwischen den ermittelten Faktoren erster Ordnung ermittelt werden (AWANG, 2015).

Zur Überprüfung der Annahme, dass ein Konstrukt (hier Systemkompetenz) aus mehreren Subkonstrukten (hier: vier Dimensionen) besteht, eignet sich als einfachste Form von Strukturgleichungsmodellen die konfirmatorische Faktorenanalyse zweiter Ordnung (Second Order CFA), bei welcher neben den vier Dimensionen als Faktoren erster Ordnung zusätzlich das Gesamtkonstrukt Systemkompetenz sys als übergeordneter Faktor zweiter Ordnung definiert wird (vgl. [ABB. 3](#)).

[ABB. 3](#) zeigt, dass sich die unstandardisierten Faktorladungen der Items der vier Kompetenzdimensionen gegenüber dem Modell erster Ordnung nur leicht verändern. Der entscheidende Unterschied ergibt sich darin, dass die latenten Variablen der vier Systemkompetenzdimension nun nicht mehr miteinander kovariieren. Stattdessen weist der übergeordnete Faktor zweiter Ordnung sys Faktorladungen zweiter Ordnung auf die Subkonstrukte auf, d.h. die Kovarianzen gehen in diesen übergeordneten Faktor auf. Vereinfacht ausgedrückt wird die Gemeinsamkeit der vier Faktoren durch den übergeordneten Faktor zweiter Ordnung

abgebildet. Auch hier wird die Faktorladung wieder beim zuerst berücksichtigten Faktor (Dimension A) auf 1 gesetzt (gestrichelter Pfeil) und für alle weiteren Faktoren (durchgezogene Pfeile) in Relation hierzu berechnet.

Zusammenfassend zeigt [ABB. 3](#), dass der Faktor zweiter Ordnung sys am meisten Varianz in Dimension B aufklären kann, dicht gefolgt von Dimensionen D und C, während in Dimension A deutlich weniger Varianzaufklärung durch den übergeordneten Faktor erfolgt. Die Bedeutung der vier zugeordneten Systemkompetenzdimensionen für das Gesamtkonstrukt Systemkompetenz ist folglich entgegen der zunächst getroffenen Annahme nicht identisch, sondern weist die dargestellte Schwerpunktsetzung auf.

Bezugnehmend auf die dargestellte didaktische Auseinandersetzung mit dem Konstrukt Systemkompetenz und seiner Dimensionalität liegt ein entscheidender Schritt der vorliegenden Studie darin, die Modellpassung des aufgestellten Second Order Modells zu überprüfen. Bei einem schlechten Modellfit wäre die theoretische Annahme, dass die vier bestätigten Faktoren (Systemkompetenzdimensionen) tatsächlich gemeinsam zum übergeordneten Gesamtkonstrukt Systemkompetenz beitragen, anhand des vorliegenden Datensatzes statistisch nicht haltbar. [TAB. 2](#) zeigt daher das First Order sowie Second Order Modell im Vergleich.

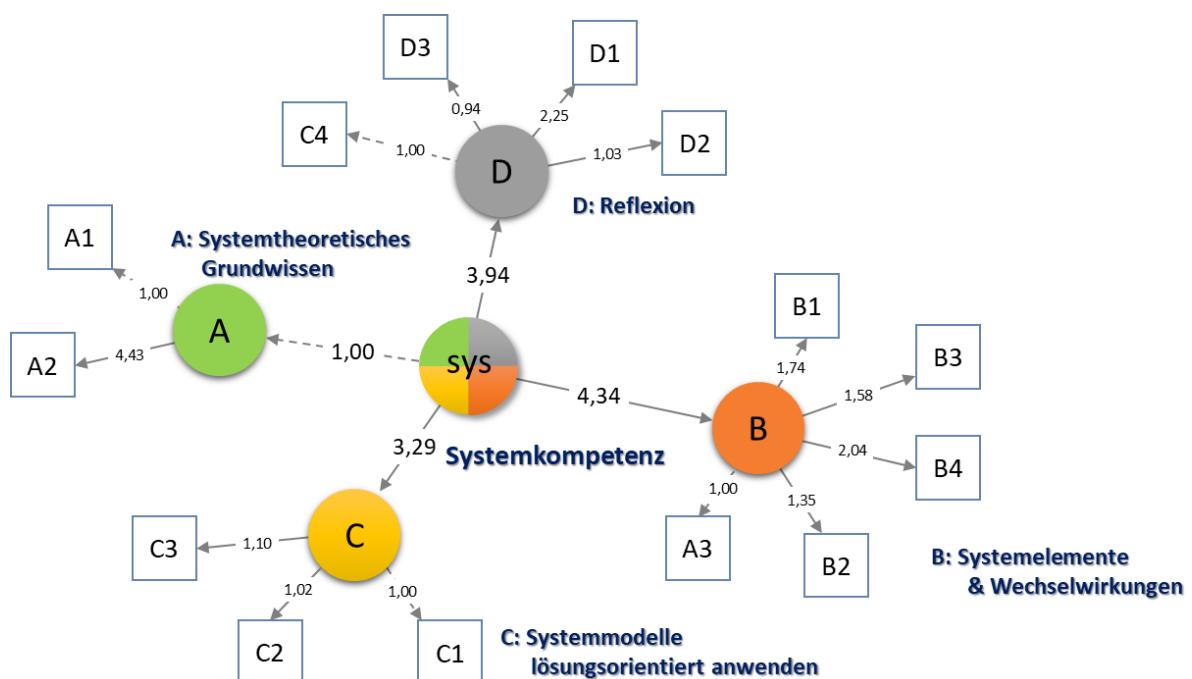


ABB. 3 Pfaddiagramm des Mess- und Strukturmodells zur Systemkompetenz (Second Order Model) (Quelle: AUTORIN & AUTOR)

Tab. 2: Modellgütekoeffizienten der Modelle erster und zweiter Ordnung für das Konstrukt Systemkompetenz (Quelle: AUTORIN & AUTOR)

Modell	Freiheitsgrade	Gesamtanpassung des Modells		Modellvergleiche		
		SRMR	RMSEA	AIC	BIC	CFI
First Order Modell 4 Faktoren (systemkompetenzdimensionen A bis D)	71	0,061	0,069	11.468	11.580	0.927
Second Order Modell wie First Order Modell zzgl. übergeordneter Faktor „sys“ (Konstrukt Systemkompetenz)	73	0,065	0,068	11.467	11.573	0.926

5.3 Diskussion und Interpretation der Faktorenstruktur

Hinsichtlich der Gesamtanpassung (SRMR; RMSEA) und auch bezüglich des CFI sind beide Modelle, wie in TAB. 2 ersichtlich, akzeptabel. Bezüglich der informationstheoretischen Maße AIC sowie BIC ist das Modell zweiter Ordnung geringfügig überlegen und wird daher bevorzugt angenommen. Das heuristisch abgeleitete vierdimensionale Freiburger Systemkompetenzstrukturmodell (RIESS et al., 2015) kann anhand des Datensatzes der vorliegenden Studie somit empirisch fundiert werden.

Die Ergebnisse der vorliegenden Studie bestätigen die beiden Dimensionen B *Systemmodellierungsfähigkeit* und C *Fähigkeit zur Nutzung von Systemmodellen beim Lösen von komplexen dynamischen Problemen* des Freiburger Systemkompetenzmodells, die ihre Entsprechungen in Dimensionen weiterer geographie-

didaktischer Kompetenzstruktur- und -stufenmodelle finden (vgl. Kap. 2.2). Darüber hinaus machen sie unter Beachtung der erreichten Modellgütemaße die zusätzliche inhaltliche Bedeutung der Theorie- sowie der Reflexionsebene für das Konstrukt Systemkompetenz im Kontext Bodenerosion deutlich, indem sich auch Dimensionen A *Deklaratives/konzeptuelles systemisches Wissen* und D *Bewertung von Systemmodellen und Ergebnissen der Modellanwendung* als eigenständige Dimensionen zeigen. Bei der Gesamtbeurteilung der statistischen Hinweise auf das vierdimensionale Modell anhand der Daten der vorliegenden Studie bleibt die im Vergleich mit den weiteren Dimensionen schwache Reliabilität der Dimension A (vgl. Kap. 4) kritisch anzumerken. Die deutlichen Unterschiede der Dimension A gegenüber den anderen Dimensionen könnten dabei neben einer schwächeren Messgenauigkeit auch auf einer schwächeren inhaltlichen Zugehörigkeit zum Konstrukt Systemkompetenz beruhen.

6 Interventionsstudie zur Förderbarkeit von Systemkompetenz

Angeschlossen an die Teilstudie zur Überprüfung der Systemkompetenzstruktur wird der entwickelte und validierte Systemkompetenztest (vgl. Kap. 4)

in der zweiten Teilstudie, einer Interventionsstudie, als Prä-Post-Messinstrument zur Evaluation von drei Lernsettings (Treatmentgruppen) eingesetzt.

6.1 Einsatz analoger, digitaler und kombinierter Modelle im Vergleich

Im Rahmen der Interventionsstudie wird die Wirksamkeit analoger bzw. digitaler Modelle bzw. einer Kombination aus beiden zur Förderung von Systemkompetenz mittels Varianzanalyse (vgl. Kap. 3) vergleichend untersucht. Bei der hier eingesetzten zweifaktoriellen gemischten Varianzanalyse mit Messwiederholung auf einem Faktor (R-Paket *ez*; LAWRENCE, 2016) werden die einzelnen Testergebnisse des Prä- bzw. Posttests vergleichend herangezogen: So sind zugleich Aussagen über die Gruppenunterschiede zwischen Prä- und Posttest sowie innerhalb von Prä- bzw. Posttest mög-

lich. Das Ergebnis lässt folglich die Interpretation von drei Effekten zu: zum einen der beiden Haupteffekte Gruppe sowie Zeitpunkt (zur Klärung der Frage, ob signifikante Unterschiede zwischen den Messzeitpunkten bzw. zwischen den Gruppen bestehen), und zum anderen des Interaktionseffekts, der im Sinne der gestellten Forschungsfrage beantworten kann, ob signifikante Gruppenunterschiede zwischen den Messzeitpunkten bestehen. Die graphische Darstellung erfolgt mittels des R-Paketes *ggplot2* (WICKHAM, 2016).

Als erstes Ergebnis kann festgehalten werden, dass der Effekt des Faktors Gruppe allein mit $p = 0,82$ statistisch nicht bedeutsam ist. Der Effekt des Faktors Zeitpunkt ist mit $p = 7,31 \cdot 10^{-45}$ hingegen signifikant und zeigt (mit großer Effektstärke Cohens f von $0,611$) an, dass die Treatments grundsätzlich förderwirksam sind: Die Systemkompetenz-Mittelwerte des zweiten Messzeitpunktes (Posttest) liegen, unabhängig von den Gruppen, insgesamt statistisch bedeutsam höher als zum ersten Messzeitpunkt (Prätest).

Der Interaktionseffekt zwischen Gruppe und Zeitpunkt ist mit $p = 0,024$ statistisch bedeutsam (Signifikanzniveau $0,05$) und weist mit Cohens f von $0,1$ eine kleine Effektstärke (nach COHEN, 1988) auf. Paarweise Vergleiche (Post-Hoc-Test der Fishers Least Significant Difference – FLSD = $4,198$) zeigen, dass die Mittelwertunterschiede zwischen Prä- und Posttest in jeder der drei Treatmentgruppen statistisch signifikant sind. Die Hypothese, dass die aktive Herangehensweise (hypothesenprüfendes Experimentieren) bei allen drei methodischen Zugängen Systemkompetenz bedeutsam fördert, kann somit bestätigt werden.

Die drei Treatmentgruppen weisen zum Messzeitpunkt 1 (Prätest) keine statistisch bedeutsamen Mittelwertunterschiede zueinander auf. Die Gruppenunterschiede im Ausgangsniveau der Systemkompetenz gehen also vor der Durchführung der drei Treatments nicht über zufallsbedingte Unterschiede hinaus. Signifikante Gruppenunterschiede zum Messzeitpunkt 2 (Posttest) lassen damit auf statistisch bedeutsame Unterschiede in der Förderwirkung der drei Treatments schließen. Hier zeigt der paarweise Vergleich, dass der beobachtete signifikante Effekt zwischen der Gruppe Kombination (= höchste durchschnittliche Gesamtscore) und der Gruppe digitales Modell (= niedrigste durchschnittliche Gesamtscore) auftritt (vgl. ABB. 4). Die Hypothese, dass die Kombination aus analogem Modell und Computersimulation bei gleicher Interventionsdauer den größten Systemkompetenzzuwachs erzielt, kann daher bestätigt werden, mit der Einschränkung, dass der Unterschied lediglich zum rein digitalen Modell statistisch signifikant ausfällt.

6.2 Systemkompetenzveränderung in den vier betrachteten Dimensionen

Im Unterschied zu der in Kapitel 6.1 dargestellten Varianzanalyse mit Messwiederholung, wird für die folgende Analyse der Zeitfaktor durch Differenzbildung zwischen Prä- und Posttestergebnissen rechnerisch entfernt und die einfaktorielle Varianzanalyse ohne Messwiederholung kommt zur Anwendung. Mithilfe

des Treatmentfaktors *Gruppe* und des errechneten Systemkompetenz-Differenzwertes in den einzelnen Kompetenzdimensionen A bis D als abhängiger Variablen lassen sich u.a. dimensionsspezifische Unterschiede in der Förderwirkung der drei Treatments anschaulich betrachten. Die bei der einfaktoriellen Varianzanalyse ohne Messwiederholung ermittelten Effektgrößen lassen sich zudem direkt mit den Effektgrößen weiterer Studien vergleichen (vgl. Eid et al., 2011, S. 453). Dies erleichtert beispielsweise den Einbezug der Ergebnisse in Metastudien.



ABB. 4 Mittelwerte der skalierten Systemkompetenz-Scores (Prä-Post) gesamt (Gesamtsäule) und nach Systemkompetenz-Dimensionen A–D (farbig) für die drei Treatmentgruppen „analoges Modell“, „digitales Modell“ und „Kombination“ (Quelle: AUTORIN & AUTOR)

Bezüglich des Systemkompetenz-Zugewinns in den Einzeldimensionen zeigt sich ein signifikanter Mittelwertunterschied nur in Dimension B, und zwar zwischen der Gruppe *digitales Modell* mit niedrigstem durchschnittlichen Kompetenzzuwachs und der Gruppe *Kombination* mit höchstem Kompetenzzugewinn zwischen Prä- und Posttest (vgl. ABB. 5). Mit Cohens f von $0,2$ handelt es sich um einen kleinen bis mittleren Effekt (nach COHEN, 1988).

In den Dimensionen A, C und D unterscheiden sich die drei Treatmentgruppen hinsichtlich der erzielten mittleren Systemkompetenzzugewinne statistisch nicht bedeutsam, jedoch lassen sich insgesamt folgende Tendenzen beobachten: In den Dimensionen A und D erzielt die Gruppe *analoges Modell* den stärksten Systemkompetenzzugewinn zwischen Prä- und Posttest, gefolgt von der Gruppe *Kombination*, während die Gruppe *digitales Modell* den jeweils niedrigsten Kompetenzzuwachs erzielt. Dieser fällt jedoch in Dimension A, in welcher bereits im Prätest gute Ergeb-

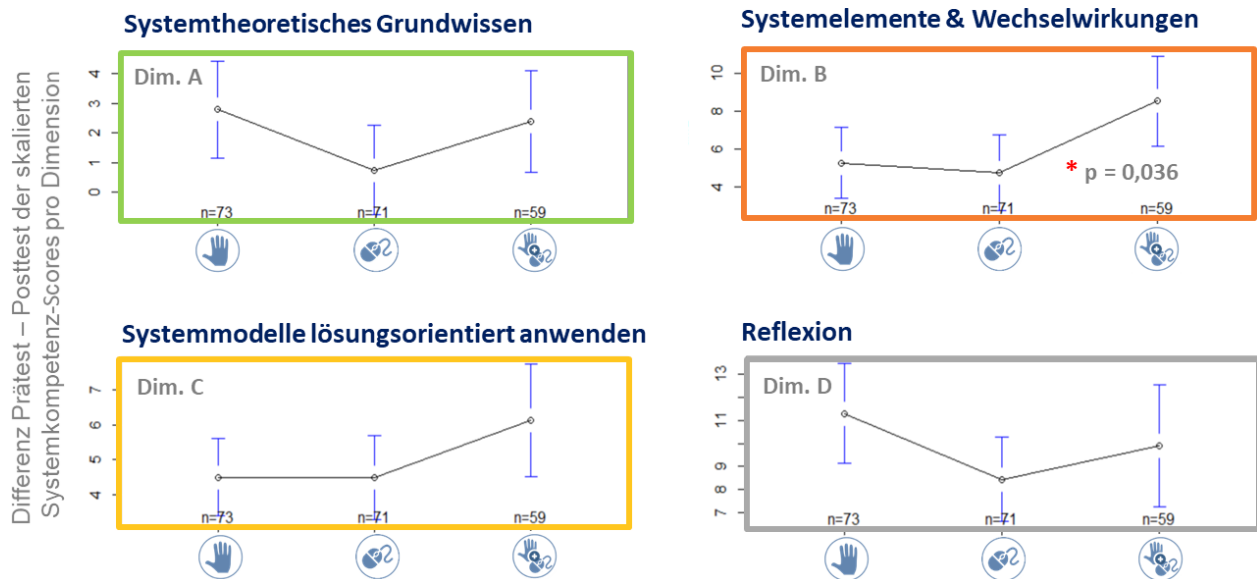


ABB. 5 Varianzanalyse ohne Messwiederholung zum Vergleich der Systemkompetenzveränderung (= Differenz zwischen Prä- und Posttest der Mittelwerte der skalierten Systemkompetenz-Scores) in den einzelnen Systemkompetenz-Dimensionen A–D

nisse erreicht wurden, insgesamt – also über alle drei Treatments hinweg – deutlich niedriger als in Dimension D aus. In Dimension C gestaltet sich das Ergebnis vergleichbar der oben beschriebenen Dimension B: die Gruppen *analoges Modell* und *digitales Modell* erzielen einen ähnlich gering ausgeprägten Systemkompetenzzuwachs, während die Gruppe *Kombination* am besten abschneidet, jedoch mit geringerem absoluten Lernzuwachs im direkten Vergleich mit Dimension B (vgl. ABB. 5).

6.3 Diskussion und Interpretation der Ergebnisse der Interventionsstudie

Bezugnehmend auf die Forschungsfrage, ob verschiedene methodische Zugänge – namentlich der Einsatz analoger, digitaler oder kombinierter Modelle – das kontextuelle systemische Denken von Schülerinnen und Schülern unterschiedlich fördern, kann aus den dargestellten Studienergebnissen am Beispiel der Erarbeitung der komplexen Dynamik von Bodenerosion gefolgert werden, dass der kombinierte Einsatz analoger und digitaler Modelle, verglichen mit dem alleinigen Einsatz digitaler Modelle, bei gleichem Zeitaufwand zu einer signifikant besseren Förderung der Systemkompetenz von Schülerinnen und Schülern führt (Gesamtbetrachtung über alle Dimensionen hinweg, vgl. ABB. 4). Hierfür wurde eine kleine Effektstärke von Cohens $f = 0,1$ ermittelt. Die angenommene Bedeutsamkeit des Effekts wurde daher bei der Stichprobenumfangsplanung überschätzt. Um eine Power von mindestens 0,80 zu erzielen, sollte bei möglichen

Folgestudien der Stichprobenumfang auf 388 Probandinnen und Probanden pro Treatmentgruppe erhöht werden. Aufgrund der Vielzahl von Faktoren, die auf das menschliche Denken und Handeln Einfluss nehmen, treten häufig keine großen Effekte eines einzelnen Faktors auf, dennoch kann er Bedeutung und inhaltliche Relevanz haben. RASCH, FRIESE, HOFMANN und NAUMANN (2014, S. 52) betonen die Notwendigkeit, die Bewertung eines Effekts ausschließlich von inhaltlichen Überlegungen abhängig zu machen. Da die Bedeutung der Systemkompetenz für die Geographiedidaktik und die Naturwissenschaftsdidaktiken im Speziellen, aber auch grundsätzlich als zukunftsfähige Denkweise für eine nachhaltige Entwicklung, zur Ausbildung von Problemlösefähigkeit sowie als Basis von Handlungs- und Gestaltungskompetenz als maßgeblich angesehen wird, werden – ihre Fördermöglichkeit betreffend – auch kleine Effekte als relevant betrachtet.

Im Folgenden werden die erzielten Ergebnisse zunächst in den in Kapitel 2.1 skizzierten Forschungsstand eingeordnet. STULL und HEGARTY (2016, S. 15) kamen im Rahmen ihrer Studie zur zusammenfassenden Feststellung, dass der Einsatz von manipulierbaren, interaktiven Modellen vielfältige Lernprozesse generell fördere, wobei es nicht darauf ankomme, ob es sich um digitale oder gegenständliche Modelle handele. Dieser Befund ist vor dem Hintergrund der vorliegenden Studie nicht zu teilen, da in Bezug auf Systemkompetenzförderung im Kontext Bodenerosion der Einsatz kombinierter Modelle dem alleinigen Einsatz digitaler Modelle signifikant überlegen ist (ABB. 4). Ungeachtet der mangelnden Vergleichbarkeit von

monokausal ausgerichtetem funktionalem Denken und multikausal ausgerichtetem Systemdenken kann die theoretisch abgeleitete Empfehlung von **LICHTI und ROTH (2018, S. 101)** des kombinierten Einsatzes analoger und digitaler Medien für den Schulunterricht für das hier untersuchte Fallbeispiel unterstrichen und empirisch abgesichert werden: Auf Basis einer expliziten Prüfung der Wirksamkeit des kombinierten analogen und digitalen Modelleinsatzes zeigt sich die Kombination dabei signifikant wirksamer als der jeweilige Einsatz nur eines Modells. Die hier gezeigten Ergebnisse bestätigen somit auch die einführend beschriebenen Beobachtungen von **RIESS und MISCHO (2008)** bezüglich der höchsten Wirksamkeit eines kombinierten Einsatzes von Computersimulation und weiteren Medien oder Methoden zur Systemkompetenzförderung, welche jedoch entweder einem didaktischen Vorteil oder der längeren Zeitdauer der kombinierten Intervention geschuldet sein könnte. Ein Mehrwert im Design der hier gezeigten Studie liegt dabei in der gleichen Zeitdauer der drei verglichenen Treatments, wodurch das signifikante Ergebnis des kombinierten Einsatzes von Modellen und Computersimulationen besondere Bedeutsamkeit erlangt.

Dieser kombinierte Modelleinsatz erzielt auf Betrachtungsebene der vier Einzeldimensionen in den beiden Dimensionen B und C die höchste Förderwirkung, während in den beiden Dimensionen A und D das analoge Modell überlegen ist (**ABB. 5**). Aufgrund der abgesehen von Dimension B nicht signifikanten Gruppenunterschiede werden im folgenden lediglich Tendenzen interpretiert. Dem Parallelisieren von Modell und Realität (Modellreflexion, Dimension D des Freiburger Modells) wird besondere unterrichtliche Bedeutung beigemessen (**KRAUTTER, 2015, S. 268; SCHUBERT, 2013, S. 199–200**). Dies gelingt in der vorliegenden Studie den am analogen Modell Lernenden am besten. Hierfür ist möglicherweise die konkretere Veranschaulichung des Prozesses Bodenerosion im analogen Modell ursächlich, während das schlechtere Abschneiden der Gruppe mit digitalem Modelleinsatz zugleich konsistent mit der Beobachtung verschiedener Autoren ist, dass hier Probleme insbesondere beim Transfer zwischen Computermodeilen und Fallbeispielen auftreten (**KLIEME & MAICHLE, 1994; BOLLMANN-ZUBERBÜHLER & KUNZ, 2008, S. 52**).

Konzeptuelles systemisches Grundwissen, wie es mit den Items der Dimension A erfasst wird, ist bei den Teilnehmerinnen und Teilnehmern der vorgestellten Studie bereits im Prätest auf einem im Durchschnitt

hohen Kompetenzlevel, mit geringer Steigerung zum Posttest (vgl. **ABB. 4**, grüne Säule). Auch wenn in den Interventionen ein Fokus auf dem praktischen systemischen Handeln am entsprechenden Bodenerosions-Modell lag, waren theoretische Reflexionen über Systeme im Allgemeinen und speziell im Kontext des betrachteten Fallbeispiels ebenso Teil der Treatments (vgl. Kap. 4), so dass als Ursache für den geringen Kompetenzzuwachs eher Deckeneffekte des Testinstruments in Frage kommen. Bei der Entwicklung zusätzlicher Items für Dimension A, die aufgrund der geringeren Reliabilität dieser Dimension für Folgestudien empfohlen wird, erscheint es daher sinnvoll, zur besseren Ausdifferenzierung zudem erhöhte Itemschwierigkeiten einzubringen. Die Erfassung und Förderung (meta-)theoretischen Grundwissens zu systemwissenschaftlichen Betrachtungsweisen oder generellen Eigenschaften komplexer Systeme wird hierbei als gewinnbringende Basis einer spiralcurricularen Entwicklung von Systemkompetenz verstanden.

Während in den Dimensionen A und D der Systemkompetenzzugewinn durch unterrichtlichen Einsatz der Modell-Kombination nicht über das durch rein analogen Modelleinsatz erreichte Maß hinausgeht, entsteht in den Dimensionen B und C ein spezifischer didaktischer Mehrwert durch den verknüpften Einsatz von analogen und digitalen Bodenerosionsmodellen: Anhand dieser beiden zentralen Kompetenzdimensionen des Freiburger Modells, die ihre Entsprechung auch in weiteren in der Geographiedidaktik etablierten Systemkompetenzstrukturmodellen finden, ist erkennbar, dass der kombinierte Modelleinsatz im Kontext Bodenerosion die Fördermöglichkeiten durch rein analoge oder rein digitale Modelle übersteigt. Eine kognitive Überlastung findet dabei offensichtlich nicht statt, stattdessen scheinen sich die positiven Lerneffekte beider Methoden zu ergänzen, wie auch **AINSWORTH und VANLABEKE (2004)** oder **RIDING und GIMLEY (1999)** den potenziellen Nutzen kombinierter Repräsentationsformen unterstreichen. Aus didaktischer Sicht steht hierbei folglich weniger eine unterschiedliche Akzentuierung analoger und digitaler Lernphasen, sondern vielmehr eine Verschaltung des Digitalen mit dem Analogen im Vordergrund, wie es auch **BAECKER (2017)** zur Förderung von Nachhaltigkeitskompetenzen fordert.

7 Fazit und Ausblick

Zusammenfassend ist festzuhalten, dass ein in Bezug auf die Gütekriterien von psychometrischen Tests überprüfbares Messinstrument zur Erfassung von Systemkompetenz von Schülerinnen und Schülern der Klassenstufe 10–12 im Kontext Bodenerosion vorgelegt werden konnte, welches – ggf. um zusätzliche Items zur verbesserten Messung der Dimension A erweitert – auch in Folgestudien eingesetzt werden kann. Dies ergänzt zudem die von BRÄUTIGAM (2014) sowie FANTA, BRÄUTIGAM, GREI und RIESS (2017) vorgelegten Systemkompetenz-Messinstrumente in ökologischen Kontexten für Schülerinnen und Schüler der Klassenstufe 6 bzw. für Lehramtsstudierende.

Mit der vorliegenden Studie ist es gelungen, einen empirischen Beitrag zur Klärung des von BOLLMANN-ZUBERBÜHLER und KUNZ (2008, S. 52) benannten Forschungsdesiderates zu der Frage zu leisten, wie sich Systemverständnis bei Schülerinnen und Schülern im Unterricht fördern lässt. Der kombinierte Einsatz analoger und digitaler Modelle zur Erarbeitung des Themas Bodenerosion führt, verglichen mit dem alleinigen Einsatz digitaler Modelle, bei gleichem Zeitaufwand zu einer signifikant besseren Entwicklung der Systemkompetenz der Lernenden. Über die Generalisierbarkeit dieser Empfehlung auf weitere Themenbeispiele ist keine eindeutige Aussage möglich, jedoch schließen SCHECKER et al. (1999, S. 17) aus ihren Untersuchungen, dass die Grundfertigkeiten des Systemdenkens sich – z.B. bei Ähnlichkeit der verwendeten Repräsentationsformen – innerhalb des Faches transferieren lassen. Im Design einer Folgestudie wäre zudem der Einbezug der hier nicht berücksichtigten Einsatz-Reihenfolge analoger und digitaler Modelle denkbar.

Neben einer quantitativen Aussage zur besten Förderwirkung auf Systemkompetenz durch den analogen, digitalen oder kombinierten Modelleinsatz ist aus didaktischer Sicht auch eine qualitative Analyse der Kompetenzentwicklung von Interesse. Im Rahmen

der durchgeführten Interventionsstudie haben alle 203 Studienteilnehmerinnen und -teilnehmer Wirkdiagramme ihres Vorwissens zu den Einflussfaktoren auf Bodenerosion erstellt und später im Zuge der Modelarbeit in ihren jeweiligen Treatmentgruppen weiterentwickelt. Von der noch ausstehenden Auswertung dieser Wirkdiagramme ist, in Ergänzung der dargestellten varianzanalytischen Ergebnisse, ein weiterer Beitrag zu einem umfassenderen Verständnis der Lernprozesse zu erwarten, die kennzeichnend für die Entwicklung von Systemkompetenz in unterschiedlichen methodisch-didaktischen Lernsettings sind. Nicht zuletzt eignet sich die integrative Betrachtung raumwirksamer Mensch-Umwelt-Beziehungen im Rahmen solcher Lernsettings zur Profilierung des Brückenfachs Geographie.

Zudem konnte die heuristischen Annahmen einer vierdimensionalen Kompetenzstruktur zum systemischen Denken nach RIESS et al. (2015) auf Grundlage des Datensatzes der vorliegenden Studie erstmalig empirisch fundiert werden. Somit tragen die Ergebnisse zum wissenschaftlichen Diskurs bezüglich der theoretischen Annahmen zur Struktur von Systemkompetenz und den ihr zugrunde liegenden Fähigkeiten bei. Für einen kompetenten, fächerübergreifenden Umgang mit den Dynamiken des globalen Wandels nehmen Systemmodelle eine bedeutende Rolle ein. Das überprüfte Freiburger Kompetenzstrukturmodell ist dazu aufgrund seiner entsprechenden Schwerpunktsetzung besonders anschlussfähig. Als Herausforderung erweisen sich hierbei insbesondere die bereits von BROCKMÜLLER (2019, S. 148–149) oder AMMONEIT et al. (2019, S. 20–21) angedeuteten Überschneidungsbereiche von Systemkompetenz und Modellierkompetenz in räumlichen Kontexten, die Ausgangspunkte weiterer Untersuchungen darstellen können.

8 Literatur

- AINSWORTH, S. & VANLABEKE, N. (2004). Multiple Forms of Dynamic Representation. *Learning and Instruction*, 14(3), 241–255. DOI: 10.1016/j.learninstruc.2004.06.002
- AMMONEIT, R., REUDENBACH, C., TUREK, A., NAUSS, T., & PETER, C. (2019). Geographische Modellierkompetenz - Modellierung von Raum konzeptualisieren. *GW-Unterricht*, 156(4), 19–29.
- AWANG, Z. (2015). *The Second Order Confirmatory Factor Analysis*. DOI: [HTTPS://DOI.ORG/10.13140/RG.2.1.1136.3049](https://doi.org/10.13140/RG.2.1.1136.3049)
- BAECKER, D. (2017). Wie verändert die Digitalisierung unser Denken und unseren Umgang mit der Welt? In R. GLÄSS & B. LEUKERT (Hg.), *Handel 4.0: Die Digitalisierung des Handels – Strategien, Technologien, Transformation* (S. 3–24). Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-662-53332-1_1
- BIRKENHAUER, J. (1997). Modelle im Geographieunterricht. *Praxis Geographie*, 27(1), 4–8.
- BOLLMANN-ZUBERBÜHLER, B. & KUNZ, P. (2008). Ist systemisches Denken lehr- und lernbar? In U. FRISCHKNECHT-TOBLER, U. NAGEL & H. J. SEYBOLD (Hg.), *Systemdenken – Wie Kinder und Jugendliche komplexe Systeme verstehen lernen* (S. 33–52). Zürich: Pestalozzianum.
- BOSSEL, H. (2004). *Systeme, Dynamik, Simulation: Modellbildung, Analyse und Simulation komplexer Systeme*. Norderstedt: Books on Demand.
- BRANDHUBER, R., AUERSWALD, K., LANG, R., MÜLLER, A., & RIPPEL, A. (2012). *ABAG interaktiv* (Version 1.0, nicht mehr verfügbar).
- BRÄUTIGAM, J. (2014). *Systemisches Denken im Kontext einer Bildung für nachhaltige Entwicklung. Konstruktion und Validierung eines Messinstruments zur Evaluation einer Unterrichtseinheit* (Dissertation).
- BROCKMÜLLER, S. (2019). *Erfassung und Entwicklung von Systemkompetenz – Empirische Befunde zu Kompetenzstruktur und Förderbarkeit durch den Einsatz analoger und digitaler Modelle im Kontext raumwirksamer Mensch-Umwelt-Beziehungen* (Dissertation).
- BROCKMÜLLER, S., VIEHRIG, K., SCHULER, C., MRAZEK, J., VOLZ, D., & SIEGMUND, A. (2016a). Enhancement of Geographical Systems Thinking Through the Use of Models. In J. LAVONEN, K. JUUTI, J. LAMPISILKÄ, A. UITTO & K. HAHN (Hg.), *Learning Science: Conceptual Understanding* (S. 158–168). Helsinki: University of Helsinki.
- BROCKMÜLLER, S., VOLZ, D., & SIEGMUND, A. (2016b). Der Einsatz experimenteller Arbeitsweisen zur Förderung geographischen Systemverständnisses bei Schüler/innen und Lehramtsstudierenden. In K.-H. OTTO (Hg.), *Geographie und naturwissenschaftliche Bildung – Der Beitrag des Faches für Schule, Lernlabor und Hochschule: Dokumentation des 21. HGD-Symposiums im März 2015 in Bochum. Geographiedidaktische Forschungen (Band 63)* (S. 104–123). Münster: Verlagshaus Monsenstein und Vannerdat.
- BÜHNER, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München etc.: Pearson.
- COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Erlbaum.
- DAHL, G. (1971). Zur Berechnung des Schwierigkeitsindex bei quantitativ abgestufter Aufgabenbewertung. *Diagnostica*, 17(3), 139–142.
- DE JONG, T., LINN, M. C. & ZACHARIA, Z. C. (2013). Physical and Virtual Laboratories in Science and Engineering Education. *Science*, 340(6130), 305–308. DOI: 10.1126/SCIENCE.1230579
- DGfG (DEUTSCHE GESELLSCHAFT FÜR GEOGRAPHIE) (Hg.) (2017). *Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss mit Aufgabenbeispielen*. Bonn: Selbstverlag DGfG.
- EDSALL, R. & WENTZ, E. (2007). Comparing Strategies for Presenting Concepts in Introductory Undergraduate Geography. *Journal of Geography in Higher Education*, 31(3), 427–444. DOI: 10/BF3R97
- EID, M., GOLLWITZER, M. & SCHMITT, M. (Hg.). (2011). *Statistik und Forschungsmethoden: Lehrbuch. Mit Online-Materialien*. Weinheim, Basel: Beltz.
- EPSKAMP, S. & STUBER, S. (2017). *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output. R package version 1.1*. Verfügbar unter <https://CRAN.R-project.org/package=semPlot>
- FANTA, D., BRÄUTIGAM, J., GREIFF, S. & RIESS, W. (2017). Entwicklung und Validierung eines Messinstruments zur Erfassung von systemischem Denken bei Lehramtsstudierenden in ökologischen Kontexten. *Zeitschrift für Didaktik der Naturwissenschaften*, 23, 241–259. DOI: 10/GD2ZVN
- FAUL, F., ERDFELDER, E., LANG, A.-G. & BUCHNER, A. (2007). G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavior Research Methods*, 39(2), 175–191. DOI: 10.3758/BF03193146

- FISSENI, H.-F. (1997). *Lehrbuch der psychologischen Diagnostik*. Göttingen, Seattle: Verlag für Psychologie.
- FLAIG, H. (2013). *Anpassungsstrategie an den Klimawandel – Fachgutachten für das Handlungsfeld Landwirtschaft*. Karlsruhe: Ministerium für Umwelt, Klima und Energiewirtschaft.
- FREIN, T., MÖLLER, G., PETERMANN, A. & WILPRICHT, M. (2007). Was sind eigentlich Pfadmodelle? Wirkungszusammenhänge aufdecken und Abhängigkeiten identifizieren. *Schulverwaltung Nordrhein-Westfalen*, 2007(10), 286–287.
- GEORGE, D. & MALLERY, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference*. Boston: Allyn and Bacon.
- HORN, J. L. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*, 30, 179–185. DOI:10.1007/BF02289447
- HU, L. & BENTLER, P. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. DOI: 10.1080/10705519909540118
- JEBBINK, K. (2013). Simulationen. In D. BÖHN & G. OBERMAIER (Hg.), *Wörterbuch der Geographiedidaktik. Begriffe von A–Z* (S. 254–255). Braunschweig: Westermann.
- KLIEME, E. & MAICHLE, U. (1994). *Modellbildung und Simulation im Unterricht der Sekundarstufe I. Auswertung von Unterrichtsversuchen mit dem Modellbildungssystem MODUS*. Bonn: Institut für Bildungsforschung.
- KÖCK, H. (2018). Fachliche Konsistenz und Spezifik in allen Dimensionen des Geographieunterrichts. In A. REMPFLE (Hg.), *Wirksamer Geographieunterricht* (S. 110–121). Baltmannsweiler: Schneider Verlag Hohengehren.
- KRAUTTER, Y. (2015). Fachtypische und überfachliche Medien im Geographieunterricht. In S. REINFRIED & H. HAUBRICH (Hg.), *Geographie unterrichten lernen. Die Didaktik der Geographie* (S. 228–271). Berlin: Cornelsen.
- LAWRENCE, M. A. (2016). *ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0*. Verfügbar unter <https://CRAN.R-project.org/package=ez>
- LICHTI, M., & ROTH, J. (2018). How to Foster Functional Thinking in Learning Environments Using Computer-Based Simulations or Real Materials. *Journal for STEM Education Research*, 1(1–2), 1–55. DOI: 10.1007/s41979-018-0007-1.
- LIENERT, G. A. & RAATZ, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz, Psychologie Verlags Union.
- MEHREN, M., MEHREN, R., OHL, U. & RESENBERGER, C. (2015). Die doppelte Komplexität geographischer Themen. Eine Herausforderung für Schüler und Lehrer. *Geographie aktuell und Schule*, 37(216), 4–11.
- MEHREN, R. (2018). Die Geographie behandelt die großen Herausforderungen des 21. Jahrhunderts. In A. REMPFLE (Hg.), *Wirksamer Geographieunterricht* (S. 130–138). Baltmannsweiler: Schneider Verlag Hohengehren GmbH.
- MEHREN, R., REMPFLE, A., BUCHHOLZ, J., HARTIG, J. & ULRICH-RIEDHAMMER, E. M. (2018). System Competence Modelling: Theoretical Foundation and Empirical Validation of a Model Involving Natural, Social and Human-Environment Systems. *Journal of Research in Science Teaching*, 55(5), 685–711. DOI: 10.1002/TEA.21436
- MEHREN, R., REMPFLE, A., ULLRICH-RIEDHAMMER, E.-M., BUCHHOLZ, J. & HARTIG, J. (2016). Systemkompetenz im Geographieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 147–163. DOI: 10.1007/s40573-016-0047-Y
- MEHREN, R., ULRICH-RIEDHAMMER, E.-M., REMPFLE, A., BUCHHOLZ, J. & HARTIG, J. (2015). Wie lässt sich Systemdenken messen. *Geographie aktuell und Schule*, 37(215), 4–15.
- MEHREN, R., REMPFLE, A. & ULRICH-RIEDHAMMER, E. M. (2014). Denken in komplexen Zusammenhängen. Systemkompetenz als Schlüssel zur Steigerung der Eigenkomplexität von Schülern. *Praxis Geographie* 44(4), 4–8.
- OHL, U. (2018). Komplexität. In A. BRUCKER, J.-B. HAVERSATH, & A. SCHÖPS (Hg.), *Geographie-Unterricht: 102 Stichworte* (S. 121–122). Baltmannsweiler: Schneider Verlag Hohengehren.
- OTTO, K.-H. (2015). Experimente. In S. REINFRIED & H. HAUBRICH (Hg.), *Geographie unterrichten lernen: Die Didaktik der Geographie* (S. 144–149). Berlin: Cornelsen.
- PAAS, F., TUOVINEN, J. E., TABBERS, H. & VAN GERVEN, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, 38(1), 63–71. DOI: 10.1207/S15326985EP3801_8

- RASCH, B., FRIESE, M., HOFMANN, W. & NAUMANN, E. (2014). *Quantitative Methoden 1: Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Berlin, Heidelberg: Springer.
- REMPFLER, A., & MEHREN, R. (2018). Systemisches Denken. In A. BRUCKER, J.-B. HAVERSATH & A. SCHÖPS (Hg.), *Geographie-Unterricht: 102 Stichworte* (S. 205–206). Baltmannsweiler: Schneider Verlag Hohengehren.
- RIDING, R. & GIMLEY, M. (1999). Cognitive Style, Gender and Learning from Multimedia Materials in 11-Year-Old Children. *British Journal of Educational Technology*, 30(1), 43–56. DOI: 10.1111/1467-8535.00089
- RIESS, W. (2013). Bildung für eine nachhaltige Entwicklung (BNE) und Förderung des systemischen Denkens. *Anliegen Natur*, 35(1), 55–64.
- RIESS, W., SCHULER, S. & HÖRSCH, C. (2015). Wie lässt sich systemisches Denken vermitteln und fördern? *Geographie aktuell und Schule*, 37(215), 16–29.
- RIESS, W., & MISCHO, C. (2008). Wirkungen variierten Unterrichts auf systemisches Denken. In U. FRISCHKNECHT-TOBLER, U. NAGEL & H. J. SEYBOLD (Hrsg.), *Systemdenken – Wie Kinder und Jugendliche komplexe Systeme verstehen lernen* (S. 135–147). Zürich: Pestalozzianum.
- ROSSEEL, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. DOI: 10.18637/jss.v048.i02
- ROST, J., LAUSTRÖER, A. & RAAK, N. (2003). Kompetenzmodelle einer Bildung für eine nachhaltige Entwicklung. *Praxis der Naturwissenschaften – Chemie in der Schule*, 52(8), 10–15.
- RUDDLE, R. & JONES, D. M. (2001). Manual and Virtual Rotation of a Three-Dimensional Object. *Journal of Educational Psychology*, 7(4), 286–296. DOI: 10.1037/1076-898X.7.4.286
- SCHECKER, H., KLIEME, E., NIEDDERER, H., EBACH, J. & GERDES, J. (1999). *Physiklernen mit Modellbildungssystemen – Förderung physikalischer Kompetenz und systemischen Denkens durch computergestützte Modellbildung (Abschlussbericht für die Deutsche Forschungsgemeinschaft DFG)*. Aufgerufen am 23. Januar 2020 unter http://www.idn.uni-bremen.de/pubs/DFG_PMS_Ab.pdf
- SCHEURING, M. & ROTH, J. (2017). Real Materials or Simulations? Searching for a Way to Foster Functional Thinking. In T. DOOLEY & G. GUEUDET (Hg.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education (CERME10, February 1 – 5, 2017)* (S. 2678–2679). Dublin: DCU Institute of Education & ERME.
- SCHUBERT, J. C. (2013). Modelle. In D. BÖHN & G. OBERMAIER (Hg.), *Wörterbuch der Geographiedidaktik. Begriffe von A–Z* (S. 199–200). Braunschweig: Westermann.
- SCHUMACKER, R. E. & LOMAX, R. G. (2004). *A Beginner's Guide to Structural Equation Modeling*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- SENGE, P. M. & KLOSTERMANN, M. (2006). *Die fünfte Disziplin: Kunst und Praxis der lernenden Organisation*. Stuttgart: Klett-Cotta.
- SEYBOLD, H., FRISCHKNECHT-TOBLER, U. & NAGEL, U. (2008). Systemisches Denken – Aufgaben für Forschung und Schulpraxis. In U. FRISCHKNECHT-TOBLER, U. NAGEL & H. J. SEYBOLD (Hg.), *Systemdenken – Wie Kinder und Jugendliche komplexe Systeme verstehen lernen* (S. 149–156). Zürich: Pestalozzianum.
- SOMMER, C. (2006). *Untersuchungen zur Systemkompetenz von Grundschulern im Bereich Biologie* (Dissertation).
- STEIGER, J. H. (2007). Understanding the Limitations of Global Fit Assessment in Structural Equation Modeling. *Personality and Individual Differences*, 42(5), 893–898. DOI: 10.1016/j.paid.2006.09.017
- STULL, A. T. & HEGARTY, M. (2016). Model Manipulation and Learning: Fostering Representational Competence with Virtual and Concrete Models. *Journal of Educational Psychology*, 108(4), 509–527. DOI: 10.1037/EDU0000077
- TRIONA, L. M. & KLAHR, D. (2003). Point and Click or Grab and Heft: Comparing the Influence of Physical and Virtual Instructional Materials on Elementary School Students' Ability to Design Experiments. *Cognition and Instruction*, 21(2), 149–173. DOI: 10.1207/S1532690XCI2102_02
- VIEHRIG, K., SIEGMUND, A., WÜSTENBERG, S., GREIFF, S., & FUNKE, J. (2017). The Heidelberg Inventory of Geographic System Competency Model. In D. LEUTNER, J. FLEISCHER, J. GRÜNKORN & E. KLIEME (Hg.), *Competence Assessment in Education: Research, Models and Instruments* (S. 31–53). Cham: Springer.
- VIEHRIG, K., SIEGMUND, A., WÜSTENBERG, S., GREIFF, S., & FUNKE, J. (2012). Systemisches und räumliches Denken in der geographischen Bildung. In A. HÜTTERMANN, P. KIRCHNER, S. SCHULER & K. DRIELING (Hg.), *Räumliche Orientierung. Räumliche Orientierung, Karten und Geoinformation im Unterricht* (S. 95–102). Braunschweig: Westermann.

WICKHAM, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer: New York.

WIKTORIN, D. (2013). Graphische Modelle im Geographieunterricht: Handlungsorientierter Einsatz von und kritischer Umgang mit Modellen. *Praxis Geographie*, 43(12), 4–7.

WINBERG, T. M. & BERG, C. A. R. (2007). Students' Cognitive Focus during a Chemistry Laboratory Exercise: Effect of a Computer-Simulated Prelab. *Journal of Research in Science Teaching*, 44(8), 1108–1133. DOI: [10.1002/TEA.20217](https://doi.org/10.1002/TEA.20217)